

Using Aesthetic Judgements to Distinguish between Humans and Computers

Nasser Mohammed Al-Fannah
Information Security Group
Royal Holloway, University of London
nasser@alfannah.com

April 12, 2017

Abstract

As a result of continuing advances in computer capabilities, it is becoming increasingly difficult to distinguish between humans and computers in the digital world. We propose using the fundamental human ability to distinguish between things that are aesthetically pleasing and those that are not as the basis of a method to verify that a communicating party is human. We discuss one possible implementation of this notion to develop a new CAPTCHA, the *Aesthetic CAPTCHA*, which we compare with widely used CAPTCHAs. Our initial analysis shows that, at least in theory, Aesthetic CAPTCHAs offer advantages over other schemes in terms of satisfying the full range of CAPTCHA requirements. More generally, using human aesthetic judgement adds a possible new dimension to the future design of Turing tests.

1 Introduction

In 1950, Turing [1] discussed how computers might demonstrate intelligence indistinguishable from human intelligence. A Turing test (or Imitation Game as described by Turing) is a test of the ability of a human (called the Evaluator) to distinguish between a computer and a human via a typed conversation and within a limited time frame. Today the term Turing test is used more loosely to describe any method for distinguishing between a human and a computer [2].

So called CAPTCHAs [3] (the term being loosely derived from Completely Automated Public Turing test to tell Computers and Humans Apart) are an important class of Turing tests in this broader sense. A CAPTCHA is a challenge-response test (or puzzle) that many websites use to try to ensure that the entity interacting with the website is human and not a computer program.

In practice, many CAPTCHAs require the user to re-type a displayed distorted text, where the distortion is intended to render the text readable by

humans but not by computers. However, many currently used CAPTCHAs are difficult for humans to use, easily solved by a computer program, or both [4].

In this paper, we propose the use of the human ability to make aesthetic judgements as the basis for an effective Turing test. This is based on the belief that making an aesthetic judgment would be difficult to program into a computer, while it is easily practiced by most humans [5]. The remainder of the paper is structured as follows. Following a brief introduction to Turing tests in section 2, CAPTCHAs are reviewed in section 3, followed by brief discussion of some widely used CAPTCHAs in section 4. A discussion of aesthetic judgements is given in section 5. A human/computer distinguishing system is introduced in section 6, followed by the proposed Aesthetic CAPTCHA in section 7. In sections 8 and 9, details of a theoretical analysis and practical trials of the Aesthetic CAPTCHA are provided. Some of the most widely used CAPTCHAs are compared with the Aesthetic CAPTCHA in section 10; concluding remarks follow in section 11.

2 Turing Tests

There appear to be two main applications of Turing tests. One involves evaluating Artificial Intelligence (AI) implementations. Tests of this type take a variety of forms, but typically depend on AI software being able to produce human-like conversations or judgements. At the present time, no AI can simulate general intelligence and so cannot pass a general Turing test (i.e. on an unlimited range of topics) [6]. Instead AIs are usually tested in the less demanding scenario of a specific problem or field. Such tests are known as subject matter expert Turing tests (SMETTs) [7]. For example, a computer (i.e. running an AI application) might attempt to imitate a human in giving responses and advice related to marketing. The software would then be evaluated on how well it has performed against human experts in the subject. The results are usually ranked based on how close they are to human performance.

The other main application of Turing tests is in the form of CAPTCHAs, also known as Human Interaction Proofs (HIPs). The main difference between the two applications of Turing tests is that, when used to test the effectiveness of AI systems, the test is typically conducted by a human; by contrast, the key property of a CAPTCHA is that it can be conducted automatically by a computer. In other words, the evaluator in an AI evaluation is a human whereas the evaluator in a CAPTCHA is a computer. Moreover, CAPTCHAs are intended to be solved by humans.

In principle, aesthetic judgements can be utilized in both applications. However, in this paper we focus only on the CAPTCHA application. We next discuss CAPTCHAs in greater depth.

3 CAPTCHAs

Unlike Turing tests used for evaluating AIs, CAPTCHAs are meant to be automated in the sense that a computer is taking the role of the evaluator in the Imitation Game instead of a human. CAPTCHAs are typically used by websites as an automated way of blocking interactions initiated by computer programs commonly known as bots. These bots are used to automatically accomplish undesirable tasks such as: creating large number of accounts (e.g. email accounts to be used for malicious purposes such as originating spam), manipulating on-line polls, or creating large volumes of traffic to a website to cause a Denial of Service (DoS) [3].

Many commonly visited websites¹ use CAPTCHAs, including YouTube, MSN, Google, Yahoo, Facebook and Twitter. However, CAPTCHAs suffer from two main practical problems. One is that in many cases it has been shown to be possible to program bots to solve a specific CAPTCHA — see, for example, [8, 9, 10]. The other is that, in practice, humans often find it challenging to solve CAPTCHAs [11].

It is generally accepted (see, for example, [11]) that any CAPTCHA should possess the following four properties.

1. *Ease for Computers to Evaluate*: the correctness of the solution to the puzzle should be simple for the computer to verify.
2. *Ease for Humans*: the puzzle should be easy for humans to solve.
3. *Ease of Generation*: the generation of puzzles in software should be straightforward.
4. *Challenging for Computers*: solving a puzzle should be difficult for a computer.

Existing CAPTCHAs, regardless of their type, often test one or more of the following human capabilities that are believed to be very challenging for computers to reproduce.

1. *Invariant recognition*: this refers to the human ability to recognize previously unseen variations of a known object, image or text.
2. *Segmentation*: this is the human ability to segment and recognize connected or overlapping characters or objects.
3. *Context*: captures the human ability to understand context, and thereby more easily recognize a distorted or vague image or text. Moreover, humans are able to detect anomalies through innate understanding of what is out of context [12].

¹The CAPTCHA Usage website <http://trends.builtwith.com/widgets/captcha> provides continuously updated information about the current use of CAPTCHAs.

4 Widely Used CAPTCHAs

We next briefly describe four very widely used CAPTCHAs. These provide the basis for a comparative analysis of the Aesthetic CAPTCHA, introduced in section 7. The CAPTCHAs we describe are image-based reCAPTCHA, text-based reCAPTCHA, No CAPTCHA reCAPTCHA (NCRC) and sweetCaptcha². According to a recent survey 90% of websites use Google’s reCAPTCHA (including NCRC), whereas only just under 1% use sweetCaptcha. It is worth noting that Google’s reCAPTCHA has three types: text-based, image-based and checkbox-based; these three variants are discussed separately below. A further CAPTCHA called “Are You a Human”³ is also widely used. However, we do not consider it further here because it is customized for individual websites through advertisements, and this lack of a specific CAPTCHA puzzle format makes it difficult, if not impossible, to compare with other CAPTCHAs.

4.1 Text-based reCAPTCHA

This CAPTCHA uses images from Google maps *street view* that contain numbers, text or both (see figure 1). An image is displayed and the user is required to type the numbers or text that appear.

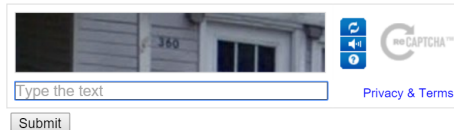


Figure 1: A Text-based reCAPTCHA puzzle

4.2 Image-based reCAPTCHA

In these puzzles, the user is shown nine images and asked to select all those that satisfy a specific condition (see figure 2). The condition could, for example, be “select all images with flowers”. The number of images that satisfy the condition is usually three.

4.3 No CAPTCHA reCAPTCHA (NCRC)

This CAPTCHA simply asks the user to check a box that states *I’m not a robot* (see figure 3). Google claims that it uses an advanced risk analysis engine to

²This selection covers four of the most widely used schemes, according to statistics found on <http://trends.builtwith.com/widgets/captcha> as of 08/04/2017; information regarding the four selected CAPTCHAs can be found at <https://www.google.com/recaptcha> and <http://sweetcaptcha.com>

³Further information can be found on the CAPTCHA’s official website <https://areyouahuman.com/>

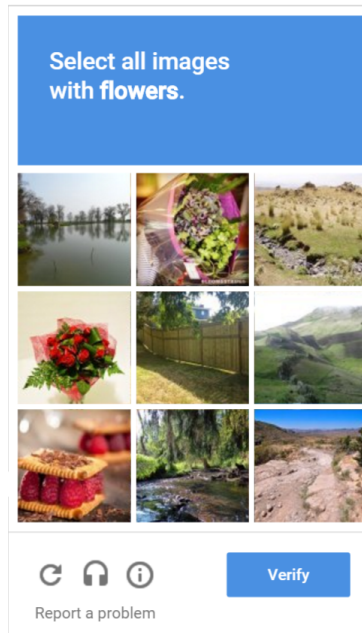


Figure 2: An Image-based reCAPTCHA puzzle

determine whether or not the user is human⁴. If, after the box has been checked, the server suspects the client is not human, it presents a reCAPTCHA puzzle (either image-based or text-based). The precise techniques used by the Google server are not public; however, informal tests reveal part of its private rule set. For example, we found that if the Google web cookies are deleted, an incognito web browser session is used, or JavaScript is disabled, then a reCAPTCHA puzzle will always be offered.

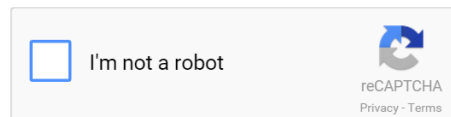


Figure 3: No CAPTCHA reCAPTCHA

4.4 SweetCaptcha

This CAPTCHA requests the user to drag the matching image from amongst four presented to a fifth image. For example, it might ask the user to *drag the correct plug to the socket* (see figure 4). The idea behind this scheme is that

⁴<https://www.google.com/recaptcha/intro/index.html>

there is a commonly known relation between the object that needs to be dragged and the object that is being dragged to.

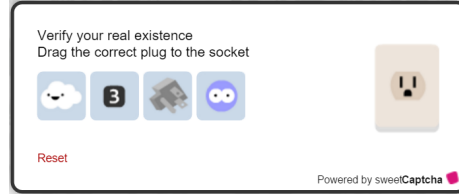


Figure 4: A sweetCaptcha puzzle

5 Aesthetic Judgements

The dictionary definition of aesthetic [13] is that it relates to what is pleasurable in appearance. It is also defined as something pleasurable to the senses through a perception of beauty. Put simply, aesthetics are related to an object, event or experience that people deem to be beautiful. Aesthetics can be related to many things, such as the sight of an object, the hearing of a sound, going through an experience, the witnessing of an event, etc.

An object, event or experience that has the property of stimulating pleasure is considered to have a positive aesthetic value. On the other hand, if it stimulates displeasure then it is considered to have a negative aesthetic value [14]. An aesthetic judgement is a human’s assessment of the aesthetic value of something. Appreciation of aesthetics is something that appears to be possessed by humans of all ages. People of all cultures decorate themselves, their possessions and surroundings for aesthetic reasons [15]. The ability to make aesthetic judgements has the advantage of being part of human nature from childhood, and so does not require a certain level of education, knowledge or language fluency. For the purposes of this paper, we restrict our attention to the hypothesis that most people make the same judgement when contrasting between objects with positive and negative aesthetic values, at least in domains of interest to us.

6 Human/Computer Aesthetic Distinguisher

We propose a general approach we call the Human/Computer Aesthetic Distinguisher. This is a system that utilizes the human ability to make aesthetic judgements as a way of distinguishing humans from computers. One reason it is difficult to program a computer to make aesthetic judgements is that the judgment process itself is not completely understood by humans. The nature of an aesthetic judgement has been, and still is, a source of controversy in terms of its universality, existence in certain respects, rating and other aspects [16].

However, our lack of complete understanding of what makes something aesthetically pleasing or not does not hinder our ability to make aesthetic judge-

ments. Despite the lack of a complete understanding of the nature of aesthetics, the field of *Computational Aesthetics* has developed computational methods that can make certain aesthetic decisions in a similar way to humans [17]. Moreover, there has been work [18] on automating the discovery of aesthetically pleasing pictures. However, to date this field has had very limited success, since it mainly depends on automatically finding professionally-taken photographs and their popularity on social media [18] rather than examining what makes a picture beautiful. This situation seems likely to continue until there is an improved understanding of the human ability to make aesthetic judgements [5].

People typically agree that a flower is prettier (and probably smells better) than roadkill. However, what if a human is presented with a number of flowers, one of which has an unpleasant smell? How good can a computer program be at making such a distinction when it is oblivious to the context and subject? We suggest that a program is likely to do rather poorly. It is worth emphasizing that to make an aesthetic judgement the subject must first recognize what they are looking at. Image recognition alone is hard for computers [19], and it is merely a prerequisite for making an aesthetic judgement. It therefore seems reasonable to assume that using aesthetics as the basis of a Turing test can make it very challenging for a computer to imitate a human.

Aesthetic judgements could involve any of the human senses. We suggest that, to make the system as simple as possible to use, subjects should not be asked to rank aesthetic values but rather simply make a judgement whether something is aesthetically pleasing or not. As well as simplifying use, such an approach is intended to ensure that agreement in the aesthetic judgement is maximized.

The human/computer aesthetic distinguisher in its most general form operates as follows.

1. The tester assembles two collections of samples (e.g. images, sound samples, etc.), of which one set consists of samples found to be aesthetically pleasing by the vast majority of human subjects and the other contains samples found to be displeasing.
2. The test subject is presented with a selection of samples from the two sets, and is asked to indicate at least one sample which is aesthetically pleasing (or displeasing).
3. The tester verifies the correctness of the sample (or samples) selected by the subject.

Clearly, constructing tests will be greatly facilitated if the assembly of samples in step 1 could to some extent be automated. Nevertheless, this may be difficult since we cannot expect a program to make aesthetic judgements! However, there may, for example, be libraries of images, all of which are aesthetically pleasing (or displeasing). Similarly, in step 2 it would help increase sample variety if the samples could be automatically modified before being presented to the subject, thereby making attempts to defeat the system by cataloguing all the

possible samples more difficult. This issue is discussed further in the following section.

7 Aesthetic CAPTCHA

In an image-based CAPTCHA, a user is typically presented with a number of images. The user might be asked to select one or more images that satisfy a specific condition (e.g. select all the images containing cats). Moreover, depending on the type of CAPTCHA, the user might be asked to perform an action, such as select, drag, align or rotate, on the image(s) that satisfy the given condition.

We now outline the Aesthetic CAPTCHA, an image-based aesthetic distinguisher in which the user is asked to select one from a number of images. To keep it simple for users while maintaining difficulty for bots, we recommend that a single user puzzle should contain 9 to 12 images. This is consistent with the current practices of major CAPTCHA providers such as Google and Microsoft⁵, as exemplified by the example reCAPTCHA puzzle shown in figure 2. When presented with these images, the user is asked to select the image that has a positive aesthetic value (i.e. is beautiful) or alternatively select the image that has a negative aesthetic value (i.e. is ugly). The images could be drawings or digital art, as well as photos.

Figure 5 illustrates a prototype of the Aesthetic CAPTCHA. The user is presented with a set of nine images of different objects of which only one is aesthetically displeasing (i.e. the rotten apple at the top centre). The user is given an instruction as simple as *click on the image that does not look nice*. Since only an aesthetic judgement is required, the puzzle instructions do not need to specify what is presented in the images. Moreover, as in the example in figure 5, the presented images do not need to all be in a single category (e.g. buildings or cars); they can be of a virtually unlimited range of objects as long as one stands out as aesthetically displeasing. Obviously such a CAPTCHA can be reversed to present eight aesthetically displeasing images with only one that is aesthetically pleasing.

It is easy to see that aesthetic judgements are relative. A certain image might be regarded as the most aesthetically displeasing in a given set but not in another. For example, some observers might consider an old classic car to be ugly, whereas others might think it beautiful. However, neither party would consider it to be the most aesthetically displeasing when contrasted with a wrecked car. In addition, a building could be aesthetically pleasing in one photo and aesthetically displeasing in another as a result of the way in which it is photographed.

As in the second example above, not only can images be chosen on the basis that they show intrinsically aesthetically pleasing or displeasing subjects, but selection can also be made depending on the way in which subjects are shown (e.g. how they are photographed) and the condition of the subjects at the time

⁵As seen in Google’s image-based reCAPTCHA and Microsoft’s discontinued Assira CAPTCHA



Figure 5: An Aesthetic CAPTCHA puzzle

of being photographed (e.g. burning, collapsing, dirty, etc.). This increases the number of options that are available for choosing images for such a CAPTCHA. Nevertheless, it means that a certain image can sometimes be suitable for use in a puzzle, and sometimes not, depending on the set of images presented with it.

8 Theoretical Analysis

Suppose a puzzle involves presenting to a user a total of n images of which k are correct (i.e. they meet a specified condition), where the user is asked to select all the k correct images ($n = 9$ and $k = 1$ in figure 5). For the aesthetic CAPTCHA, “correct” might mean pleasing or displeasing. If the user makes a random selection of k of the presented images, then the probability P of success is $P = 1/\binom{n}{k}$, where $\binom{n}{k}$ is the binomial coefficient representing the number of ways of choosing k objects from a set of n . This clearly represents a lower bound on the probability of success for puzzle-solving software.

That is, for the example in figure 5, $P = 1/9$, i.e. puzzle-solving software will have at least an 11% chance of solving a puzzle. This probability could clearly be reduced by increasing n and/or k , and/or by asking the user to solve multiple puzzles. Of course, there are two dangers with so doing, namely increasing the degree of inconvenience for the user and also increasing the failure rate for genuine users. The ability to achieve an appropriate balance between the desire to minimise the probability of correct solution by software with usability is perhaps the acid test for any CAPTCHA, and is arguably something that many current CAPTCHAs fail to achieve. More generally, all CAPTCHAs can be made more robust against automated solvers, e.g. by detecting mouse movements to detect a human presence.

A further general issue for any such image-based CAPTCHA, which applies equally to reCAPTCHA, concerns the number of images available for use in puzzles. Suppose the Aesthetic CAPTCHA is implemented using a fixed database of m images, of which p are aesthetically pleasing and $d = m - p$ are displeasing. If m , p and d are not all large, then it may be feasible for an attacker to rapidly reproduce (almost) the entire database, e.g. by having a small team of humans solve large numbers of puzzles. This argues in favour of both having a large set of images and also continually updating the database, including by automatically transforming the images to make matching harder.

Finally, we observe that the Aesthetic CAPTCHA may be unusable, or at least challenging, for users with certain disabilities. This is an expected disadvantage of such an image-based CAPTCHA, especially for people with vision disabilities. So, as with many existing CAPTCHAs, an alternative should be offered for people for whom the CAPTCHA presents difficulties. A commonly used alternative is an audio CAPTCHA, in which the user is presented with spoken words or numbers and then required to type in what they have heard [20].

9 Practical Trials

We next report on a small-scale practical trial of the Aesthetic CAPTCHA. The purpose of this trial was to check that the approach viable, i.e. that individuals can agree on which of a set of samples is aesthetically pleasing (or displeasing). To collect images for testing purposes, we used a simple Google image search query to find images at the Wikimedia Commons website⁶, a non-profit online repository of free-to-use images.

When selecting images, we avoided those which we intuitively judged that users would find hard to categorise as aesthetically pleasing or not. So, for example, when searching for flower images, we ignored unusual looking flowers or flowers filled with thorns. It took less than 15 minutes to assemble a test database of more than 200 images of flowers, cars, animals, buildings and Lego models. Thus, we estimate that 12 hours of human labour can produce approximately 10,000 images for use in an Aesthetic CAPTCHA, which theoretically can be used to generate more than one thousand non-overlapping puzzles.

For testing purposes, we classified the images we collected based on the type of object depicted (i.e. flowers, cars and buildings) and then divided each type into two groups, one containing aesthetically pleasing images and the other aesthetically displeasing images. The images included photos, paintings and hand or digital drawings. Moreover, a good number of images included other objects (e.g. people, landscape, etc.) but not as the main focus. The photos included those that seemed to have been taken by professionals as well as by amateurs.

We created a simple web-based demo of the Aesthetic CAPTCHA (see figure 5) that displays nine randomly selected images in each puzzle. Informally, the image that was aesthetically displeasing in each puzzle we generated seemed

⁶<http://www.wikimedia.org>

very easy to determine. We also saw no reason to suspect that the generated CAPTCHAs would be challenging for a human with normal vision.

To test the validity of these observations we showed these sample CAPTCHAs to 30 fairly arbitrarily selected test subjects with varying education levels and ages (including the elderly). We told the subjects that they should select the image that did not look nice compared to the rest. Each subject was asked to solve three or four different puzzles. All of the subjects were able to point out the aesthetically displeasing image in the puzzles that were presented to them in an average time of under 5 seconds.

Of course, this is far from a full scientific trial; however, it gives some confidence to our belief that the Aesthetic CAPTCHA would be simple to use. That being said, it would be undoubtedly desirable to test the scheme on a larger sample and under more scientifically controlled circumstance. It is interesting to note that, in our small-scale trial, subjects were able to solve the puzzles more quickly (on average in less than 4 seconds) if all the puzzle images were of the same type, e.g. all of cars. This is a topic worth investigating in a future larger scale trial; if this observation proves to be robust, then this not only suggests Aesthetic CAPTCHAs should be designed to use a homogeneous set of images, but it also may have analogous lessons for the design of other visual CAPTCHAs.

10 Comparisons

In this section we consider how well the Aesthetic CAPTCHA meets the four fundamental requirements for a CAPTCHA introduced in section 3, including comparisons with the four widely used CAPTCHAs described in section 4.

10.1 Assessment Criteria

We compare how well each CAPTCHA meets the requirements given in section 3. However, we omit the *ease for computers to evaluate* requirement, because four of the five CAPTCHAs simply involve comparing the user input with the stored answer (see 10.1.1).

In addition, we also narrow our consideration of the *not easily solved by bots* requirement. How well this requirement is met depends on the state of the art in the development of computer techniques to solve particular puzzles. It is particularly difficult to assess the aesthetics CAPTCHA in this respect since it has not yet been considered by the academic community. We therefore restrict our attention to something we can measure, namely the *probability of random guess success*, i.e. the probability that a random solution to a puzzle will be deemed correct.

The three criteria that are used in our comparison are therefore: ease for humans, ease of generation, and probability of random guess success. The results of the comparisons between the CAPTCHAs against the three criteria are given in 10.1.1-10.1.3. For each comparison criterion, we rank the five CAPTCHAs in

order, “1” representing the CAPTCHA that best meets the criterion. Note that this ranking is subjective for all criteria, with the exception of the probability of random guess success. Moreover, the ranking does not indicate the degree of quantitative difference between one CAPTCHA and another. The comparisons below are summarised in Table 1.

10.1.1 Ease for Humans

To obtain a preliminary understanding of the relative ease with which the various types of puzzle can be completed, we performed a small-scale, informal experiment. Confirming the results from these preliminary tests will clearly require a larger trial conducted in a more rigorous way.

30 participants were presented with 3 example puzzles for each of the types of CAPTCHA being compared, and were then asked to rank the five types in order of ease of solution. 27 of the 30 participants ranked the ease of the CAPTCHAs in the following order, from easiest to most difficult to solve: NCRC, sweetCaptcha, Aesthetic CAPTCHA, image-based reCAPTCHA and text-based reCAPTCHA.

This result is consistent with the following simple analysis.

- We first observe that it is perhaps unsurprising that text-based reCAPTCHA ranked last; this is consistent with the findings of other authors [4]. In addition, the fact that NCRC puzzles were consistently ranked the easiest to solve is unsurprising since they simply involve clicking on a checkbox.
- Image-based reCAPTCHA and Aesthetic CAPTCHA have some similarities, as they both typically present nine images to the user. However, an image-based reCAPTCHA is likely to be more difficult to solve than an Aesthetic CAPTCHA, since several images have to be selected (not just one), and, unlike Aesthetic CAPTCHA, the image selection criterion changes from one puzzle to the next.
- Since a sweetCAPTCHA only involves dragging one of four images, it is likely to be easier to solve than going through the nine images displayed in an Aesthetic CAPTCHA or image-based reCAPTCHA.

However, while NCRC appears to be the easiest to use, it is not without shortcomings, namely its dependence on cookies and JavaScript. As described in section [refncrc](#), if the appropriate cookie is not available or JavaScript is disabled, then it reverts to a “classic” reCAPTCHA, typically an image-based puzzle. Hence it does not replace existing CAPTCHAs; it merely reduces the number of occasions when one needs to be used.

10.1.2 Ease of Generation

It seems safe to assume that an NCRC instance is the easiest to generate as what is displayed is fixed, although the downloaded JavaScript may take time

Table 1: CAPTCHAs Comparison

Criteria / CAPTCHA	Text-based re-CAPTCHA	Image-based reCAPTCHA	NCRC	sweetCaptcha	Aesthetic CAPTCHA
Ease for Humans	5	4	1	2	3
Ease of generation	2	4	1	5	3
Probability of random guess success	1 (<1%)	2 (1.8%)	5 (N/A)	4 (25%)	3 (11.1%)

to execute on the client platform. On the other hand, sweetCaptcha is likely to be the most difficult to generate as it uses custom images of objects, for which there is a logical relation between the dragged and dragged-to objects (e.g. drag the nest to the bird). We also assume that text-based reCAPTCHAs are easier to generate than either of the image-based CAPTCHAs, since both image-based reCAPTCHA and Aesthetic CAPTCHA require handpicked images (as opposed to simple text). Finally, we believe that Aesthetic CAPTCHA instances are easier to generate than image-based reCAPTCHAs, since image-based reCAPTCHA requires all the images to be of the same type whereas Aesthetic CAPTCHA allows much greater freedom in mixing images in a single puzzle.

10.1.3 Probability of Random Guess Success

The third criterion we use to assess the CAPTCHAs is the probability of success of a bot that simply makes random guesses for the solution of a CAPTCHA. We implemented a naïve attack of this type using a simple macro recorder program. These are widely available programs that can record user actions such as keystrokes and mouse clicks and then automatically repeat them. We tested this attack successfully on all the CAPTCHAs in our comparison study by recording macros of an appropriate set of mouse clicks (or drags) and executing them automatically. To perform the recorded actions on a CAPTCHA, we used the *Form Filler*⁷ autofill Chrome browser add-on, which is designed to complete web forms with user-supplied data. This attack is completed with the use of an autofill browser add-on that would fill-in forms using information as per user defined criteria.

The following observations can be made about the success probability for this strategy.

- Firstly, since NCRC requires a fixed action (checking a box) the success probability is 100
- Text-based reCAPTCHA would require random guessing a string of text or numbers which would result in a negligible (<1%) probability of a

⁷Can be found on <https://chrome.google.com/webstore/detail/form-filler/bnjngeaknajbdcgpfkgnonkmifih>

random guess to be correct.

- sweetCaptcha has only 4 images to select from and would result in the highest correct guess probability (25%).
- Image-based reCAPTCHA and Aesthetic CAPTCHA percentages are 1.8% and 11.1% respectively. The calculation of image-based reCAPTCHA is based on probability of selecting 3 correct images from a set of 8^8 . It is important to note here that the correct answer is usually 3 images; however through our trials we found that it is designed to become more complicated if the puzzle is answered wrongly a few times. In such case the CAPTCHA would increase the number of correct answers as well as increasing the total number of images in a set.

The probability of random guess can be drastically reduced if some CAPTCHA safeguards are put in place. Safeguards could include detecting typing, mouse movement accuracy, etc. [?]. reCAPTCHA seems to use some safeguards since the CAPTCHA puzzles tend to become more complicated with repeated attack attempts. However, the deployed safeguards were not enough to protect against new attack described in [10]. Attack attempts on sweetCaptcha did not show any signs of safeguards use.

10.2 Summary

Table 1 provides a summary of comparisons between existing CAPTCHAs and the Aesthetic CAPTCHA. The numbers (1 to 5) in the table are used to rank the performance of the CAPTCHAs for each of the three comparison criteria.

Table 1 shows that most CAPTCHAs are amongst the best of the five for some criteria but amongst the worst for others. In particular, Aesthetic CAPTCHA is neither the best nor the worst performing for any of the criteria. Unsurprisingly, NCRC appears to perform best of the five schemes we examined; however, as we observed in section 10.1.1, because it sometimes reverts to a more traditional CAPTCHAs it can be regarded as complementing existing CAPTCHA schemes rather than replacing them.

11 Conclusion

Kurzweil asserts that the best way to build machine intelligence is to first understand human intelligence [21]. This might be considered an argument in favour of the use of aesthetic judgements in Turing tests since it is a human intelligence capability that is not fully understood. Moreover, some argue that computers can never write a moving poem, draw an artistic painting or compose a symphony. However, unlike aesthetic judgement, most humans share the latter incompetence with computers. Aesthetic judgement might be usable as

$${}_8P = \frac{(n-k)!k!}{n!} = \frac{1}{56} \approx 1.8\%$$

the basis for a new SMETT AI evaluation. Moreover, as demonstrated in this paper, it can be used as the basis for an image-based CAPTCHA.

It has been reported that Google is working on developing AI techniques that can automatically label (or caption) any image [22]. For example, software might soon be widely available that can automatically label an image as a group of people in shopping mall. This means that we can expect that it will not be long before a bot can be created to recognize the contents of images potentially breaking all conventional image-based CAPTCHAs. Using the aesthetic element introduces a new problem whose solution is beyond the capabilities of today's algorithms, and unless there is a major breakthrough in the field of computational aesthetics this is likely to remain the case.

Acknowledgements

I would like to thank Professor Chris Mitchell for his guidance, encouragement and advice.

References

- [1] Alan M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.
- [2] J.H. Moor. *The Turing Test: The Elusive Standard of Artificial Intelligence*. Studies in Cognitive Systems. Springer Netherlands, 2012.
- [3] Luis von Ahn, Manuel Blum, and John Langford. Telling humans and computers apart automatically. *Communications of the ACM*, 47(2):56–60, February 2004.
- [4] Elie Bursztein, Steven Bethard, Celine Fabry, John C. Mitchell, and Dan Jurafsky. How good are humans at solving CAPTCHAs? A large scale evaluation. In *Proceedings of the 2010 IEEE Symposium on Security and Privacy (SP '10)*, pages 399–413, Washington, DC, USA, 2010. IEEE Computer Society.
- [5] Philip Galanter. *Computers and Creativity*, chapter Computational Aesthetic Evaluation: Past and Future, pages 255–293. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [6] Edward A. Feigenbaum. Some challenges and grand challenges for computational intelligence. *Journal of the ACM*, 50(1):32–40, 2003.
- [7] E. Rich. *Artificial Intelligence*. McGraw-Hill, 3 edition, 1983.
- [8] Jeff Yan and Ahmad Salah El Ahmad. Breaking visual CAPTCHAs with naive pattern recognition algorithms. In *23rd Annual Computer Security Applications Conference (ACSAC 2007), December 10-14, 2007, Miami Beach, Florida, USA*, pages 279–291. IEEE Computer Society, 2007.

- [9] Greg Mori and Jitendra Malik. Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, CVPR '03, pages 134–141, Washington, DC, USA, 2003. IEEE Computer Society.
- [10] Suphannee Sivakorn, Iasonas Polakis, and Angelos D. Keromytis. I am robot: (deep) learning to break semantic image captchas. In *IEEE European Symposium on Security and Privacy, EuroS&P 2016, Saarbrücken, Germany, March 21-24, 2016*, pages 388–403. IEEE, 2016.
- [11] Marti Motoyama, Kirill Levchenko, Chris Kanich, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. Re: CAPTCHAs: Understanding CAPTCHA-solving services in an economic context. In *Proceedings of the 19th USENIX Conference on Security*, USENIX Security '10, pages 28–28, Berkeley, CA, USA, 2010. USENIX Association.
- [12] Oleg Starostenko, Claudia Cruz-Perez, Fernando Uceda-Ponga, and Vicente Alarcon-Aquino. Breaking text-based CAPTCHAs with variable word and character orientation. *Pattern Recognition*, 48(4):1101–1112, April 2015.
- [13] *Oxford Dictionary of English*. Oxford University Press, 2012.
- [14] Yuriko Saito. The aesthetics of unscenic nature. *The Journal of Aesthetics and Art Criticism*, 56(2):101–111, 1998.
- [15] Rolf Reber, Norbert Schwarz, and Piotr Winkielman. Processing fluency and aesthetic pleasure: Is beauty in the perceiver’s processing experience? *Personality and Social Psychology Review*, 8(4):364–382, 2004.
- [16] Andrew Lothian. Landscape and the philosophy of aesthetics: is landscape quality inherent in the landscape or in the eye of the beholder? *Landscape and Urban Planning*, 44(4):177–198, 1999.
- [17] Florian Hoenig. Defining computational aesthetics. In *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Computational Aesthetics’05, pages 13–18, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association.
- [18] Rossano Schifanella, Miriam Redi, and Luca Maria Aiello. An image is worth more than a thousand favorites: Surfacing the hidden beauty of flickr pictures. In Meeyoung Cha, Cecilia Mascolo, and Christian Sandvig, editors, *Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, May 26-29, 2015*, pages 397–406. AAAI Press, 2015.
- [19] Jurgen Schmidhuber. Multi-column deep neural networks for image classification. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pages 3642–3649, Washington, DC, USA, 2012. IEEE Computer Society.

- [20] Jennifer Tam, Jiri Simsa, Sean Hyde, and Luis von Ahn. Breaking audio CAPTCHAs. In *Advances in Neural Information Processing Systems 21, Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 8-11, 2008*, pages 1625–1632. Curran Associates, Inc., 2008.
- [21] R. Kurzweil. *The Singularity is Near: When Humans Transcend Biology*. Viking, New York, 2006.
- [22] David Glazer. Atoms, bits, and cells. *Applied & Translational Genomics*, 6:11–14, 2015.